**REAL DATA**

50 data sets mostly from the UCI repository (portion shown):

<table>
<thead>
<tr>
<th>Name</th>
<th>( n )</th>
<th>( d )</th>
<th>Outliers (%)</th>
<th>Distance Assumption Violation coefficient</th>
<th>CC statistic</th>
<th>ICA statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>6.9</td>
<td>0.58</td>
<td>-0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>ion</td>
<td>320</td>
<td>9</td>
<td>15</td>
<td>0.46</td>
<td>-0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>wave</td>
<td>600</td>
<td>19</td>
<td>10</td>
<td>0.12</td>
<td>-0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>kidney</td>
<td>2200</td>
<td>7</td>
<td>15</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>penile</td>
<td>154</td>
<td>5</td>
<td>10</td>
<td>0.12</td>
<td>0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>gisette</td>
<td>298</td>
<td>6</td>
<td>10</td>
<td>0.25</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>glass</td>
<td>2000</td>
<td>10</td>
<td>15</td>
<td>0.46</td>
<td>-0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>6.9</td>
<td>0.58</td>
<td>-0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>ion</td>
<td>320</td>
<td>9</td>
<td>15</td>
<td>0.46</td>
<td>-0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>wave</td>
<td>600</td>
<td>19</td>
<td>10</td>
<td>0.12</td>
<td>-0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>kidney</td>
<td>2200</td>
<td>7</td>
<td>15</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>penile</td>
<td>154</td>
<td>5</td>
<td>10</td>
<td>0.12</td>
<td>0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>gisette</td>
<td>298</td>
<td>6</td>
<td>10</td>
<td>0.25</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>glass</td>
<td>2000</td>
<td>10</td>
<td>15</td>
<td>0.46</td>
<td>-0.22</td>
<td>0.23</td>
</tr>
</tbody>
</table>

**EMERGENCE**

**HUBS INFLUENCE**

**DEFINITION**

\( N_k(x) \), the number of \( k \)-occurrences of point \( x \), is the number of times a point occurs among \( k \) nearest neighbors of all other points in a data set.

Hubs are points with unexpectedly high \( N_k(x) \) emerging from a skewed distribution of \( k \)-occurrences.

**REFERENCES**


**REAL DATA**

50 data sets mostly from the UCI repository (portion shown):

<table>
<thead>
<tr>
<th>Name</th>
<th>( n )</th>
<th>( d )</th>
<th>Outliers (%)</th>
<th>Distance Assumption Violation coefficient</th>
<th>CC statistic</th>
<th>ICA statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>6.9</td>
<td>0.58</td>
<td>-0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>ion</td>
<td>320</td>
<td>9</td>
<td>15</td>
<td>0.46</td>
<td>-0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>wave</td>
<td>600</td>
<td>19</td>
<td>10</td>
<td>0.12</td>
<td>-0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>kidney</td>
<td>2200</td>
<td>7</td>
<td>15</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>penile</td>
<td>154</td>
<td>5</td>
<td>10</td>
<td>0.12</td>
<td>0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>gisette</td>
<td>298</td>
<td>6</td>
<td>10</td>
<td>0.25</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>glass</td>
<td>2000</td>
<td>10</td>
<td>15</td>
<td>0.46</td>
<td>-0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>6.9</td>
<td>0.58</td>
<td>-0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>ion</td>
<td>320</td>
<td>9</td>
<td>15</td>
<td>0.46</td>
<td>-0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>wave</td>
<td>600</td>
<td>19</td>
<td>10</td>
<td>0.12</td>
<td>-0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>kidney</td>
<td>2200</td>
<td>7</td>
<td>15</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>penile</td>
<td>154</td>
<td>5</td>
<td>10</td>
<td>0.12</td>
<td>0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>gisette</td>
<td>298</td>
<td>6</td>
<td>10</td>
<td>0.25</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>glass</td>
<td>2000</td>
<td>10</td>
<td>15</td>
<td>0.46</td>
<td>-0.22</td>
<td>0.23</td>
</tr>
</tbody>
</table>

**EMERGENCE**

**HUBS INFLUENCE**

**DEFINITION**

\( N_k(x) \), the number of \( k \)-occurrences of point \( x \), is the number of times a point occurs among \( k \) nearest neighbors of all other points in a data set.

Hubs are points with unexpectedly high \( N_k(x) \) emerging from a skewed distribution of \( k \)-occurrences.

**REFERENCES**


**REAL DATA**

50 data sets mostly from the UCI repository (portion shown):

<table>
<thead>
<tr>
<th>Name</th>
<th>( n )</th>
<th>( d )</th>
<th>Outliers (%)</th>
<th>Distance Assumption Violation coefficient</th>
<th>CC statistic</th>
<th>ICA statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>6.9</td>
<td>0.58</td>
<td>-0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>ion</td>
<td>320</td>
<td>9</td>
<td>15</td>
<td>0.46</td>
<td>-0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>wave</td>
<td>600</td>
<td>19</td>
<td>10</td>
<td>0.12</td>
<td>-0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>kidney</td>
<td>2200</td>
<td>7</td>
<td>15</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>penile</td>
<td>154</td>
<td>5</td>
<td>10</td>
<td>0.12</td>
<td>0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>gisette</td>
<td>298</td>
<td>6</td>
<td>10</td>
<td>0.25</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>glass</td>
<td>2000</td>
<td>10</td>
<td>15</td>
<td>0.46</td>
<td>-0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>wine</td>
<td>178</td>
<td>13</td>
<td>6.9</td>
<td>0.58</td>
<td>-0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>ion</td>
<td>320</td>
<td>9</td>
<td>15</td>
<td>0.46</td>
<td>-0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>wave</td>
<td>600</td>
<td>19</td>
<td>10</td>
<td>0.12</td>
<td>-0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>kidney</td>
<td>2200</td>
<td>7</td>
<td>15</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>penile</td>
<td>154</td>
<td>5</td>
<td>10</td>
<td>0.12</td>
<td>0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>gisette</td>
<td>298</td>
<td>6</td>
<td>10</td>
<td>0.25</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>glass</td>
<td>2000</td>
<td>10</td>
<td>15</td>
<td>0.46</td>
<td>-0.22</td>
<td>0.23</td>
</tr>
</tbody>
</table>